Memo

To:	CCB
From:	Hans Israelsson, Mike Skov, Jeff Stevens
Date:	March 12, 1998
Subject:	PIDC 6.0: Post-analysis magnitude estimation
CC:	Bill Farrell, Ben Kohl, Keith McLaughlin, Walter Nagy, Darrin Wahl

Abstract

Changes proposed here are specific to estimation of magnitude calculation outside of *GA* and *ARS*, i.e., magnitude estimation that follows analyst review. Configuration changes are proposed to accommodate major upgrades of the database schema and related modifications of the amplitude and magnitude calculation programs (*maxsurf* and *EvLoc)* and scripts controlling surface wave processing. In addition some changes are proposed to enhance functionality and performance, including: maximum likelihood calculation of mb, more efficient surface wave processing, and improved pipeline processing of surface wave data and Ms calculation.

Statement of Objective

PIDC 6.0 is a full release containing all development and maintenance since April, 1997 on the application software of time-series monitoring system, data services and utilities of the prototype IDC, PIDC. The PIDC 6.0 proposal has been divided into seven volumes following the structure of the software configuration model by Farrell (1997). Being one of these seven volumes, the objective of the present proposal is limited to implementation in PIDC operations the following components of the Computer Software Configuration Items (CSCI) of the PIDC 6.0:

- CSCI 1 (Automatic processing)
    - Automatic detection and measurement of surface waves, LR.
    - Estimation of average surface wave magnitude, Ms.
    - Measurement of LR noise.
    - Estimation of maximum likelihood, MLE, Ms.
    - Estimation of average body wave magnitude, mb.
    - Measurement of P noise.
    - Estimation of mb MLE.

    - Estimation of weighted local magnitude, ML.
- CSCI 2 (scripts to control Ms estimates).

Summary of Proposed Changes

Infrastructure changes that are common to all of PIDC 6.0 are described by Beall et al. (1998) MacRitchie, (1997), Nagy (1997), and Sereno (1996). Appendix A enumerates the software modules covered by this memo: *DFX, maxsurf, EvLoc, ARS, analyst_log, and LPcompile.* The appendix also lists other PIDC computer software components.

Changes proposed here that are specific to estimation of magnitude calculation outside of *GA* and *ARS*. Apart from accommodating major upgrades of the database schema with new and revised tables and changes in the structure of configuration files, some proposed changes also enhance functionality and performance.

Changes in maxsurf and its configuration

The updated version of *maxsurf*, version 3.0, includes the following changes:

- Modification of the database output to write to the **parrival** and **amplitude** tables, and removal of the option to write to the **arrivalamp** table. This is necessary because of changes in the PIDC 6.0 database structure. The version of *maxsurf* currently in use, version 2.4, will not work with the new database structure.
- Addition of the “net_list” option which allows all data of a single type (long period or broadband) to be processed with a single *maxsurf* execution. This will significantly reduce the number of database connections required for surface wave processing.
- Station and network names are automatically resolved; this ensures that arrivals are recorded in the database with the appropriate station name.
- Addition of an option to read dispersion curves in binary format. The binary files can be read much faster and is a significant savings for detailed dispersion models.
- A separate small program *LPcompile* that allows conversion of dispersion curves from ASCII to binary form is delivered together with *maxsurf*.
- Addition of an option to use a regionalized phase velocity model for beamforming instead of a fixed phase velocity. Such regionalized models for arrays can be developed in the future. For the time being a fixed beam velocity (3.5 km/s) is still used.
- An improved regionalized dispersion model (from Stevens and McLaughlin, 1997). However, the performance of this new dispersion model has to be validated and approved in a *separate* CCB, before being implemented in the pipeline operations.

An updated man page for *maxsurf* is in Appendix B, and a more detailed description of the changes in the program and the associated changes in configuration parameters are given in Appendix C.

Routine recall processing for maxsurf and EvLoc Ms

With the introduction of *maxsurf* version 3.0 and PIDC 6.0, automated Ms processing will be placed under the control of a Tuxedo-managed, Distributed Application Control System (DACS). Specifically, Ms processing will be controlled as part of the routine recall processing pipeline. Control under DACS is expected to result in a more reliable operation of *maxsurf*.

Recall processing is initiated when an analyst marks an interval complete from the *analyst_log* application; currently *maxsurf* is run on an event by event basis.

Perl scripts for maxsurf and Ms calculations

Perl scripts that control *maxsurf* and *EvLoc* Ms calculations have been revised and renamed. Furthermore, a new script (*MsInterval*) has been added to support the requirement of operating the Ms pipeline from a Tuxedo managed DACS. Thus the complete Ms processing now consists of the following three Perl 5 scripts:

- *MsInterval* (new, handles converting from interval-based DACS to orid-based *maxsurf* *MsOrid*)
- *MsOrid* (formerly *ms-orid*)
- *MsConflict* (formerly *msclean*)

MsInterval is called directly from the Tuxshell controlling Ms processing. *MsInterval* utilizes getpar.pl Perl library to provide a standard par interface. Appendix D provides additional details on the revised and new Perl scripts.

Configuration of DFX-recall and DFX-noiseamp for mb noise amplitudes

The manner in which *DFX*-recall is integrated into the recall pipeline has been changed. In particular, *DFX* will now be run out of the Tuxedo-based DACS as a child of *tuxshell*. In the past under the ISIS based DACS, each time interval was processed by 6 *DFX*-recall instances running in parallel. It is proposed to run a single instance of *DFX*-recall for each interval, thereby simplifying the recall pipeline. The command line for *DFX* remains unchanged.

The *DFX*-noiseamp computation is a new addition to the recall pipeline. A single instance of *DFX* is run with a command line specifying the *DFX*-noiseamp par file. It is run after the *DFX*-recall computations and before the *EvLoc* computations in the workflow sequence of the recall pipeline. The technical descriptions of the computations performed out of *DFX*-noiseamp are described in the release notes.

The noise amplitudes for mb MLE estimates are calculated in the same way as for P signals (“amptype”= A5/2) with one exception; the time window starts 3 seconds prior to the calculated arrival time and ends 9 seconds after the calculated arrival time (/nmrd/ops/net/idc/static/DFXti/noiseamp-ti.par), whereas for P signals the window limits are 0.5 and 5.5 seconds.

Configuration of EvLoc for mb, Ms, and ML calculations

Appendix E defines the configurations of *EvLoc* for calculation of station magnitudes and average and MLE network magnitudes for mb and Ms and weighted network magnitudes for ML. Average and weighted magnitudes are based on both primary and auxiliary stations and MLE are based on primary stations only. Bootstrapping is applied to estimate uncertainties of MLE estimates (uncertainties for upperbounds are not estimated). Uncertainties of ML network magnitudes are calculated as the standard deviation of the average value, whereas currently a weighted formula is used. For consistent naming, average magnitudes for mb and Ms are assigned the magtypes “mb_ave” and “ms_ave”, compared to the current “mb” and “ms_ave”.

Expected Benefits

General benefits of PIDC 6.0 are described by Beall et al. (1998). Specific benefits to magnitude computation with regard to increased functionality and enhanced performance are:

Increased Functionality

- Maximum likelihood mb and Ms magnitudes will be a significant new feature of the PIDC REB (format GSE 2.1 is expected to be available with a ‘magnitude block’ that allows incorporation of any number of magnitudes); routine calculation of MLE magnitudes for global networks have not been carried out before; routine calculation is expected to contribute to more reliable assessments of network capabilities and to more accurate magnitudes of low magnitude events.

Owing to the Tuxedo managed pipeline for Ms processing:

- Ms processing status is recorded in the **interval** table and maybe visually monitored using *WorkFlow.*
- Failures within Ms processing pipeline can be detected and logged.

Enhanced performance:

The *maxsurf* upgrades should result in:

- Reduced processing time of LR waves because of binary files for dispersion curves.
- Reduced processing time of *maxsurf* as it now requires only two executions for a given orid.
- More reliable resolution of station and network names.

The Tuxedo managed Ms processing offers several advantages:

- The proposed time based as opposed to the current origin based Ms processing conforms with normal pipeline operations.
- Processing guaranteed, there is no risk of an analyst to fail to initiate Ms processing
- Re-queuing of failed Ms processing intervals.

Possible Risks and Dependencies

In general, the extensive changes to configuration and database structures could cause configuration errors that in turn could generate various kinds of errors. The risks for such occurrences are minimized by clear definitions and instructions for configuration changes and close monitoring of processed results that is planned once changes have been implemented.

MLE magnitudes based on one or more signal amplitudes and based on noise amplitudes only (upper bound) are currently labeled with the same magtype, which does not allow GSE 2.1 to make a distinction. If such a distinction is desirable, the magtype field will have to be updated once MLE calculations have been completed.

Noise amplitudes, in particular for LR, may occasionally be underestimated, which could result in anomalously low MLE estimates. As MLE estimates will be available through GSE 2.1, review of MLE values will add to the bulletin quality control tasks.

Initiation of Ms estimation from within *ARS* is no longer required, nor is this feature supported. Under the normal operational environment, Ms processing will be initiated as part of recall processing, which is in turn initiated by analyst interaction with the *analyst_log* program. In the testbed environment, recall processing was initiated by the publication of the REB. It is not expected that this implementation inconsistency will result in difficulties for the operational installation. Under the current operational configuration, the recall pipeline is initiated by the *analyst_log* program; Ms processing will simply be added to this existing pipeline.

The current implementation of the daily clean up script does not support repeated execution. That is, it can not be run against the same data-day more than once. The new implementation (*MsConflict*) also requires this restriction. Removing this restriction would require the creation of new database tables to store state information prior to cleaning. Modifications to the *MsConflict* script would also be required.

Because Ms processing is now under the control of Tuxedo, Ms processing should not be initiated manually. That is, MsOrid should not be executed from a UNIX shell; it should only be executed under the control of the recall pipeline. An automated routine to re-queue events for processing is not included with this release. However, intervals may be re-queued by manipulating the interval table. Depending on the circumstances, it may also be necessary to delete **stamag** and **netmag** records for the event in question.

Although the monitoring capability of the new implementation far exceeds the capability currently available to Operations, there are limitations. Once an interval is successfully processed, the *Workflow* display will not reflect changes to the events within that interval. If an analyst modifies an event, after declaring the interval complete, the event may not be properly processed.

The interval Ms processing introduces a delay as it is no longer initiated when an analyst finishes a particular event, but rather when the analyst finishes an assigned time period. This is not expected to cause any delays in the REB, as the clean-up of possible Ms mis-associations sets the ultimate time limit, and this clean-up will, as currently, be carried out after all origins of a data day have been reviewed.

The additional processing of P noise amplitude data for mb MLE adds to the total processing time of recall processing. However, experience with processing on the testbed suggests that this will not cause any significant delays.

Summary of Testing

Detection of surface waves and Ms estimation

Testing of Ms estimation was accomplished by utilizing the testbed's specially configured recall Pipeline. This pipeline processes events published in the REB, making comparisons straightforward. Values computed in Operations were compared to those computed using the new configuration. Deviations between Operations and the Testbed can occur. Missing data, for example, could cause differences between the two environments.

Ms network averages were computed for each event over a five day period, spanning 1998030 - 1998034. During this time period, 164 events were published. Of these 164 events, 18 deviations were found between Operations and the Testbed. Nine events were found to have Ms estimates differing by at least 0.01 magnitude units. These are listed in the Appendix F.

Of the 164 events, Operations published Ms estimates for 44 events, while the Testbed determined Ms for just 43 events. The Testbed process determined Ms for four specific events which did not have Ms in the Operations reviewed bulletin. These events are listed in Appendix F. While the exact reason for this deviation is not known, two strong possibilities are that either the Ms value was removed by Bulletin QC staff, or the event was un-intentionally not processed in Operations.

Five events with computed Ms values were found in the REB for which Ms was not computed by the testbed processing. All five anomalies can be attributed to failures of the recall processing pipeline. Under normal operational procedures, this failure would have been detected and displayed on the *Workflow* and corrected by Operation's personnel. Details on the five events are given in Appendix F.

P noise measurements and mb estimation

The mb MLE magnitude calculations on the testbed with the new amplitude schema and updates of *EvLoc* and associated configuration changes were evaluated using processed data for about 10 days on the testbed.

The testbed mb MLE calculations were driven by REB origins and included the following steps:

- Copy over of signal amplitudes (detection amplitudes) of detecting stations and conversion of **arrivalamp** table to the new **amplitude** table (the conversion will not be carried out during pipeline processing).
- Calculation with *DFX* of signal amplitudes of analyst-added phases.
- Calculation of noise amplitudes (event amplitudes) for non-detecting stations with *DFX*

- Calculation of mb (based on Primary and Auxiliary station data) and mb_mle (based on Primary station data only) with *EvLoc* in two separate runs.

The magnitude calculations were validated by applying algorithms independent of those in *EvLoc* to data for Jan 14, 1998. Network - average, MLE, and upper-bound (noise only) - station magnitudes agreed to within 0.01 magnitude unit for 42 events. Uncertainties for network average magnitudes had similar agreement; no independent calculation was made for uncertainties based on boostrapping for MLE magnitudes.

In addition, magnitudes calculated during the period Jan 7 - Jan 18, 1998 were reviewed. During the time period, there were minor changes in configuration parameters, for example networks used to calculate average and MLE estimates. Differences relative to proposed configuration were accounted for in the review.

Processing for one or more hours were usually missed for each data day because of various changes on the testbed. Multiple amplitudes and associated multiple station magnitudes were also noted for 8 events. These could be associated with power outages. It is assumed that data processed before an outage were still in the queue for processing as power came back and were therefore re-processed.

Comparison of average and MLE magnitudes were limited to events without amplitude duplicates and to events for which at least 90% of primary stations within 20-100 degrees in operation (based on detections in idcdel.arrival) that were processed and contributing with detection or event amplitudes. Data for about 30% of the events in the REB with mb were thus reviewed for the period. Differences between average and MLE were comparable with previous evaluations (Appendix G and H) with few outlying data points. Nine out of 203 events (or 4%) of the events had differences larger than 0.5 magnitude units. The largest difference, 0.74, occurred for events with the average magnitude based on only one station.

Schedule and Plan for Implementation

This proposal should be implemented as soon as possible after implementation of the automatic processing (Beall et al., 1998) and interactive processing components of PIDC 6.0. The following steps should be followed:

At a high level, the implementation of the post-analysis magnitude estimation routines is relatively straightforward. The details, however, are quite complex and beyond the scope of this document. Implementation will require careful review of the release notes provided with PIDC 6.0. It is necessary to implement these features simultaneously with the rest of PIDC 6.0 as the older versions of this software are not compatible with the new database schema and operate only under ISIS, which is no longer supported.

The configuration consists of a TI, several *tuxshells*, three key Perl scripts, parfiles for *EvLoc* and *maxsurf* and modifications for monitoring Ms processing via the **interval** table. It is assumed the database schema changes will be completed prior to implementation of the recall processing pipeline.

TI-recall will be required for this process, and may require minor configuration changes from the current installation.

The *tuxshells* are configured to operate the recall pipeline in a prescribed order.

1. *DFX-recall*
2. *DFX-noiseamp*
3. *EvLoc-recall*
4. *EvLoc-mlppn*
5. *EvLoc-mle*
6. *MsOrid*

Each of the *tuxshells* correspond to a Tuxedo queue which must be created by the Tuxedo Administrator. In addition to the queues, the Tuxedo Administrator will also need to configure *Scheduler* and *TMQForward*.

Additional information on *tuxshell* par files, Perl scripts, and par files for *DFX*, *maxsurf*, and *EvLoc* is given in Appendix I, which also gives an installation check-list.

Finally, *maxsurf* and *LPcompile* must be released for CMR software management configuration system and built locally.

Costs and Resources Required for Implementation

It is estimated that about one man day of Operations staff time will be required for implementation of this proposal. No costs will be incurred.

References

Beall, G., Al-Rawi, L., Fox, W., Le Bras R., Nagy, W., O'Brien M, Wahl, D., Wuester, J. and J, Zhang, PIDC 6.0: Automatic processing of seismic and hydroacoustic data, Proposal to the PIDC Configuration Control Board, CCB-PRO-98/06, 1998.

Farrell, W., Issue 2 of the PIDC Software Architecture Description, Informal Memo, Jul 30, 1997.

MacRitchie, B, Amplitude / Magnitude Schema Upgrade, Informal memo, November 10, 1997.

Nagy, W., 'Amplitude Schema Upgrade Process and Flow Design.', Informal memo, Sep 10, 1997.

Rezapour, M. and R. G. Pearce, 'Bias and Data Censoring in ISC Surface Wave Magnitudes', Bull. Seism. Soc. Am., 88, 43-61, 1998.

Sereno, T., 'Amplitudes and Magnitudes at the IDC', Informal Memo, May, 1996.

Stevens, J. L. and K. L. McLaughlin, “Improved Methods for Regionalized Surface Wave Analysis,” Maxwell Technologies Final Report submitted to Phillips Laboratory, MFD-TR-97-15887, September, 1997.